Whitman
SCHOOL of MANAGEMENT
SYRACUSE UNIVERSITY

MBC 638
DATA ANALYSIS AND DECISION MAKING

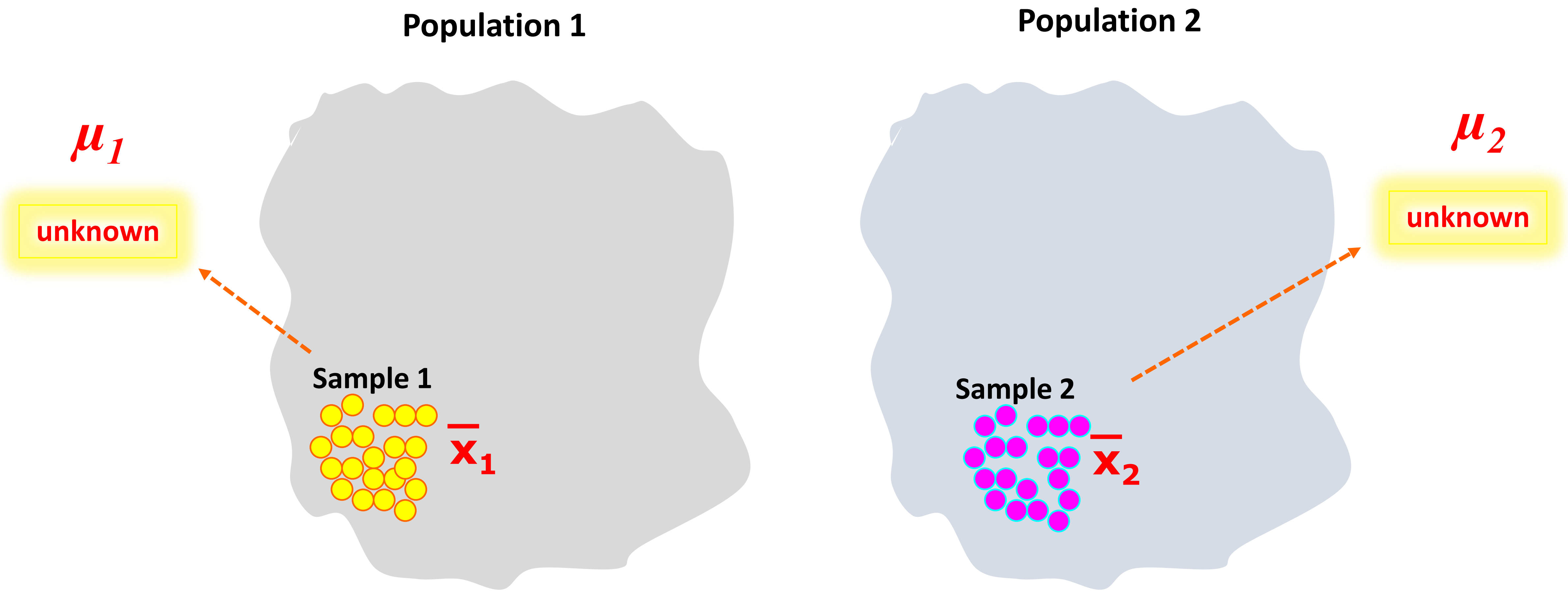
Anna Chernobai, Department of Finance

HYPOTHESIS TESTING for the difference in population means
Chapter 9

Today:

Hypothesis testing for $\mu_1 - \mu_2$

- We use information from 2 observed samples to conclude how μ_1 and μ_2 compare.
- <u>Assumption</u>: The two samples are **independent**. (We will relax this assumption later.)



Example

μ_1

We want to test hypothesis that, on average, data scientists that have a graduate degree earn a higher salary than those that don't have a graduate degree.

μ_2

Hypothesized mean difference under the null hypothesis

$$H_0: \quad \mu_1 \leq \mu_2$$

$$H_A: \quad \mu_1 > \mu_2$$

It's the same as: $H_0: \quad \mu_1 - \mu_2 \leq 0$

It's the same as: $H_A: \quad \mu_1 - \mu_2 > 0$

Recall: In hypothesis testing for μ

Test statistic is the **Z-score** of the observed **sample mean**

$$\frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

Analogously, in hypothesis testing for $\mu_1 - \mu_2$

Test statistic is the **Z-score** of the **difference** between the observed **sample means**

$$\frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Hypothesized mean difference under the null hypothesis

Left-tail test

$$H_0: \quad \mu_1 \geq \mu_2 \quad (i.e. \quad \mu_1 - \mu_2 \geq 0)$$
$$H_A: \quad \mu_1 < \mu_2 \quad (i.e. \quad \mu_1 - \mu_2 < 0)$$

$$p - value = P[\, z \quad < \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}\,]$$

Right-tail test

$$H_0: \quad \mu_1 \leq \mu_2 \quad (i.e. \quad \mu_1 - \mu_2 \leq 0)$$
$$H_A: \quad \mu_1 > \mu_2 \quad (i.e. \quad \mu_1 - \mu_2 > 0)$$

$$p - value = P[\, z \quad > \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}\,]$$

2-tail test

$$H_0: \quad \mu_1 = \mu_2 \quad (i.e. \quad \mu_1 - \mu_2 = 0)$$
$$H_A: \quad \mu_1 \neq \mu_2 \quad (i.e. \quad \mu_1 - \mu_2 \neq 0)$$

$$p - value = 2 \bullet P[\, z \quad > \left| \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \right|\,]$$

If σ_1 and σ_2 are unknown, use ***t***-distribution (or Z approximation if n_1 and n_2 are both ≥ 30)

Example

Periodically, Merrill Lynch customers are asked to evaluate Merrill Lynch financial consultants and services (*2000 Merrill Lynch Client Satisfaction Survey*). Higher ratings on the client satisfaction survey indicate better service, with 7 the maximum service rating.

Independent samples of service ratings for two financial consultants—Ms. Chen and Mr. Wang are summarized here. Ms. Chen has 10 years of experience, whereas Mr. Wang has 1 year of experience.

	Ms. Chen (10 yrs)	Mr. Wang (1 yr)
# reviews collected	36	30
Mean	6.58	6.25
Standard deviation	0.64	0.75

a) At α=5%, determine whether consultants with more experience get **higher** average service rating.

b) At α=5%, determine whether consultants with the two different levels of experience get **different** average service ratings.

What if the samples are **not independent**?

Paired samples

Example Sales Presentation Ratings.xlsx

Example Task Times.xlsx

Today: Hypothesis testing for $\mu_1 - \mu_2$

- Independent samples ✓
- Dependent (paired) samples ✓